

Doctors and Clinicians Star Ratings on Medicare Care Compare 2020 Public Reporting

Overview

The Centers for Medicare & Medicaid Services (CMS) is publicly reporting 2020 Quality Payment Program (QPP) performance information on [Medicare Care Compare](#) and in the [Provider Data Catalog \(PDC\)](#). A subset of 2020 Merit-based Incentive Payment System (MIPS) quality and Promoting Interoperability measures are reported using star ratings on clinician and group profile pages and in the PDC.

CMS finalized an item-level benchmark as the basis for clinician and group star ratings in the 2016 Physician Fee Schedule (PFS) Final Rule ([80 FR 71128 through 71129](#)). Star ratings are only publicly reported if the performance information meets the established public reporting standards and resonates with users (§414.1395(b)). The first star ratings were publicly reported in late 2017 for a subset of group-level Physician Quality Reporting System (PQRS) performance information and will continue to be publicly reported each year forward ([82 FR 53829](#)).

Visit the [Care Compare: Doctors and Clinicians Initiative page](#) to download the 2020 Clinician and Group Star Rating Cut-offs.

Select one of the topics below to learn more:

- [Why a benchmark?](#)
- [The ABC™ methodology](#)
 - [How will the benchmark be calculated?](#)
- [What about star ratings?](#)
 - [Equal ranges method](#)
- [More ways to learn](#)

Why a benchmark?

Benchmarks are important for ensuring that patients and caregivers accurately interpret and understand performance information. They allow patients and caregivers to best understand performance information by setting a point of comparison and providing context.

*Benchmarks help us **interpret and understand** performance information by setting a point of comparison.*

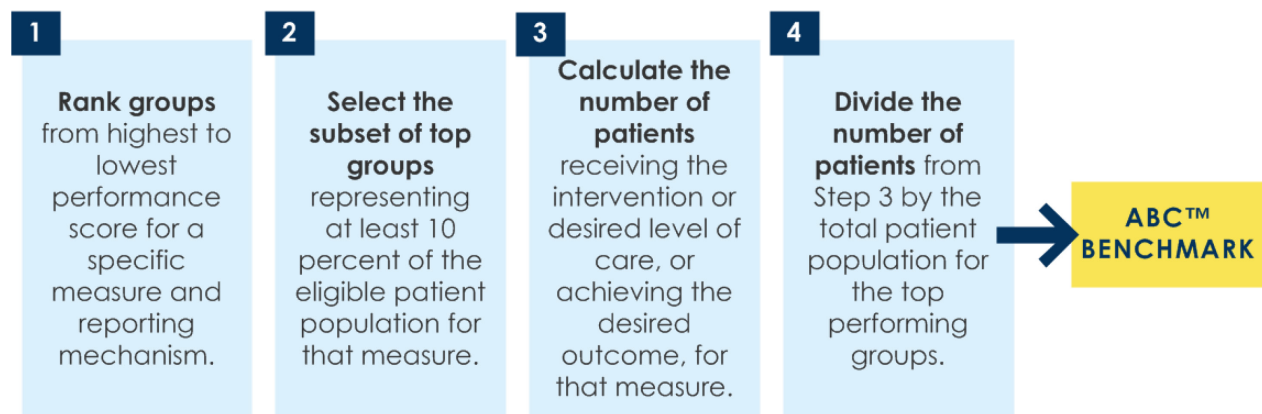
The ABC™ methodology

The Achievable Benchmark of Care (ABC™) methodology is used to develop the benchmarks that anchor star ratings for doctors and clinicians. The use of the benchmark was first finalized in the Calendar Year (CY) 2016 PFS Final Rule ([80 FR 71128 through 71129](#)). ABC™ is a well-tested, data-driven methodology. It represents quality while being realistic and achievable. It also encourages continuous quality improvement and is shown to lead to improved quality of care.^{1,2,3}

How is each benchmark calculated?

An ABC™ is established by reporting entity and collection type. ABC™ starts with the pared-mean. This is the average of the best performers on a measure for at least 10% of the patient population – not the population of clinicians or groups reporting on the measure. Figure 1 provides a step-by-step breakdown of how the benchmark is calculated.

Figure 1. Benchmark Calculation



¹ Kiefe CI, Weissman NW, Allison JJ, Farmer R, Weaver M, Williams OD. Identifying achievable benchmarks of care: Concepts and methodology. *International Journal of Quality Health Care*. 1998 Oct; 10(5):443–7.

² Kiefe CI, Allison JJ, Williams O, Person SD, Weaver MT, Weissman NW. Improving Quality Improvement Using Achievable Benchmarks for Physician Feedback: A Randomized Controlled Trial. *JAMA*. 2001; 285(22):2871–2879.

³ Wessell AM, Liszka HA, Nietert PJ, Jenkins RG, Nemeth LS, Ornstein S. Achievable benchmarks of care for primary care quality indicators in a practice-based research network. *American Journal of Medical Quality* 2008 Jan–Feb; 23(1):39–46.

We first rank-order reporters from highest to lowest performance score. Next, we include a calculation of a beta binomial model adjustment to account for low denominators. This ensures that very small sample sizes do not over-influence the benchmark but still allow all data to be included in the benchmark calculation. Then, we create a subset of the reporters by selecting the best performers until we have selected enough reporters to represent at least 10% of all patients relevant for that measure.

We establish a benchmark by calculating mean performance across these top performers. This produces a benchmark that represents the best care provided to at least 10% of patients. For a benchmark to be calculated, the measures must meet our public reporting standards (§414.1395(b)). Each measure must prove to be statistically accurate, valid, and reliable. Also, the measure must prove to resonate with patients and caregivers via testing. If these criteria are met, then we calculate the benchmark. The benchmark itself must also meet our statistical reporting standards.

What about star ratings?

Star ratings are a user-friendly way to share complex information. Star ratings give patients and caregivers more context to best understand performance information. For example, on some measures, a clinician's raw score of 80% is considered very good relative to other clinicians' performances on that measure at this moment in time. However, without star ratings, users may not realize this and assume 80% is just average performance. Star ratings help patients and caregivers accurately evaluate performance scores because these ratings provide a point of comparison.

After we determine the benchmark for a given measure, reporters that meet or exceed the benchmark are assigned 5 stars. The next step in moving to star ratings was to decide on a method for assigning 1 to 4 stars. As requested by stakeholders and encouraged by the many experts consulted, we focused on a method that:

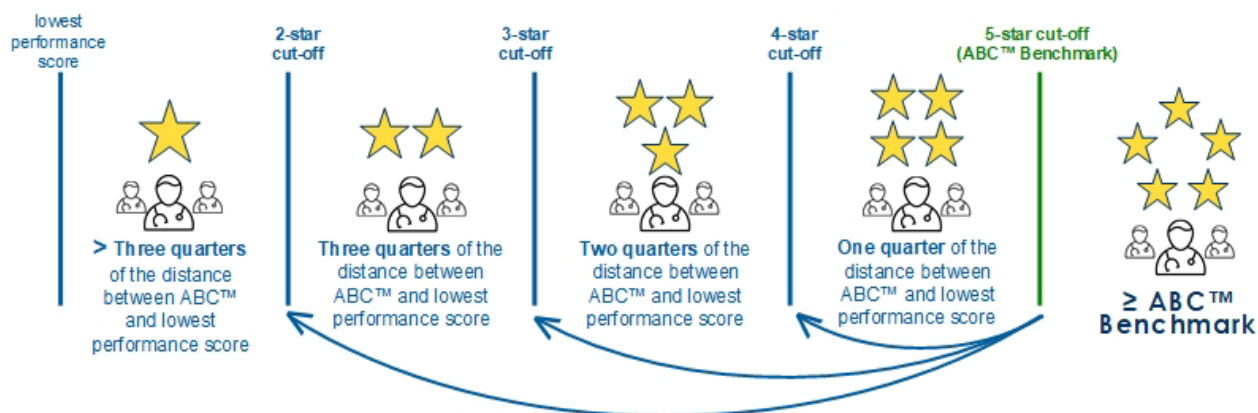
1. Avoids forcing a star-rating distribution;
2. Does not make it hard to achieve a moderate to good rating; and
3. Reliably assigns reporters into a star rating.

As discussed in the CY 2018 QPP Final Rule ([82 FR 53827 through 53829](#)) and CY 2019 QPP Final Rule ([83 FR 59915](#)), we conducted extensive statistical analyses, sought expert input, and reached out to stakeholders, including specialty societies and professional groups, to help determine the best possible method for assigning 1 to 4 stars. This work led to a decision to use the equal ranges method to assign star ratings, starting with a subset of group-level 2016 PQRS performance scores that were publicly reported in 2017. The same approach is used to assign star ratings to a subset of QPP performance information.

Equal ranges method

The equal ranges method is based on the difference between the ABC™ benchmark and the lowest performance score⁴ for a given measure. The method uses that range to assign 1 to 4 stars. Reporters that meet or exceed the established ABC™ benchmark for a measure will be assigned 5 stars.

Figure 2. Equal Ranges Method



To determine the 4-star cut-off, we subtract the lowest performance score from the ABC™ benchmark to get the range of performance scores for that measure, and then divide by 4 to give us quarters. The 4-star cut-off is one quarter of the distance between the ABC™ benchmark and the lowest performance score. Reporters that score at or above the 4-star cut-off, but below the benchmark will be assigned 4 stars.

We use the same idea to determine the 3-star cut-off. The 3-star cut-off is two quarters of the distance between the ABC™ benchmark and lowest performance score. Reporters that have scores at or above the 3-star cut-off but below the 4-star cut-off are assigned 3 stars.

We follow the same method to get the 2-star cut-off, which is 3 quarters of the distance between the ABC™ benchmark and lowest performance score. Finally, any scores that are greater than 3 quarters of the distance between the ABC™ benchmark and the lowest performance score are assigned 1 star.

More ways to learn

To learn more about public reporting and star ratings for doctors and clinicians on Care Compare, visit the [Care Compare: Doctors and Clinicians Initiative page](#). [Subscribe](#) to the Care Compare: Doctors and Clinicians listserv to receive the latest information and updates.

If you have questions, please contact the QPP Service Center at QPP@cms.hhs.gov.

⁴ To ensure that the star rating cut-offs are not overly influenced by a single performance rate, the minimum theoretical non-outlier performance score, or lower bound, used to establish the star rating cut-offs is calculated by taking the 25th percentile performance rate value and subtracting 3 times the interquartile range (IQR). If this equals less than 0%, the lower bound is set to 0%.